International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Review on Soil Classification Using Machine Learning Techniques

P Gopala Raju, M Sowmya, N Chakradhar, R Nithin Kumar

GMR Institute of Technology, 3rd year civi,Bachlor of Technology,vizianagaram,532127,India

A B S T R A C T

This review paper systematically examines the latest developments in soil classification methodologies, focusing on the integration of innovative machine-learning algorithms across diverse soil-related studies. In the classification step of soil analysis, researchers explore the efficacy of diverse machine learning algorithms to accurately categorize soil types. The investigated algorithms encompass a range of methodologies, each with its unique approach to pattern recognition and classification. Decision Trees, as a fundamental algorithm, provide a hierarchical structure for decision-making based on input features, enabling the identification of distinct soil classes. Artificial Neural Networks (ANN) leverage interconnected nodes mimicking the human brain's neural architecture, allowing them to capture intricate relationships within soil data for classification purposes. Support Vector Machines (SVM) handle classification by finding optimal hyperplanes to separate different soil classes in a feature space, which is particularly useful for complex and nonlinear datasets. Random Forest, an ensemble learning method, combines multiple Decision Trees to enhance accuracy and robustness in soil classification tasks. The review encompasses studies employing machine learning classifiers, comparing their performance in diverse contexts such as land-use land-cover (LULC) mapping. Different datasets, preprocessing techniques, and evaluation metrics are analyzed, offering insights into the strengths and limitations of each approach. The paper also covers data imputation strategies, including KNN imputation and SMOTE, addressing missing data and class imbalances in soil datasets. The diverse applications and methodologies discussed provide a comprehensive overview of the evolving landscape of soil classification, offering valuable guidance for researchers and practitioners seeking efficient and accurate soil characterization techniques.

Keywords: Soil classification, Machine Learning, Soil parameters, Algorithms, Support Vector Machines, Artificial Neural Networks

1. INTRODUCTION

Land use classification plays a pivotal role in understanding and managing the Earth's diverse landscapes, serving as a fundamental tool for sustainable resource management, urban planning, and environmental conservation. Traditional methods of land use classification often rely on remote sensing data and manual interpretation, which can be time-consuming and subjective. In recent years, there has been a paradigm shift towards leveraging advanced technologies, particularly Machine Learning (ML) techniques, to enhance the accuracy and efficiency of land use classification processes. This review paper focuses on the innovative integration of soil parameters into the realm of land use classification, exploring the synergy between soil science and ML algorithms. Soil parameters, such as texture, organic matter content, and nutrient levels, carry valuable information about the inherent characteristics of the land that influence its use and productivity. Incorporating these parameters into ML models presents a promising avenue for refining the accuracy of land use classification and providing a more holistic understanding of terrestrial ecosystems. Machine Learning, with its ability to discern complex patterns and relationships within large datasets, has demonstrated remarkable success in land use classification tasks. By training models on diverse datasets that encompass both spectral and soil information, researchers have been able to develop robust algorithms capable of differentiating between land cover types with unprecedented accuracy. The integration of soil parameters not only enhances the discriminatory power of ML models but also contributes to a more comprehensive understanding of the factors influencing land use patterns. Throughout this review, we will delve into the key ML techniques employed in land use classification, ranging from traditional algorithms like Support Vector Machines (SVM) and Random Forests to more advanced deep learning approaches like Convolutional Neural Networks (CNN). Additionally, we will explore case studies and applications where the fusion of soil parameters and ML techniques has yielded significant advancements in land use classification accuracy. As we navigate the landscape of this interdisciplinary field, the synthesis of soil science and machine learning promises to revolutionize our ability to characterize and classify land use, fostering sustainable land management practices and facilitating informed decision-making for the challenges posed by a dynamic and ever-changing environment.

2. Literature Review

The integration of soil parameters and Machine Learning (ML) techniques in land use classification has gained substantial attention in recent literature, reflecting a paradigm shift towards more robust and accurate methods. Numerous studies highlight the efficacy of ML models in land use classification, with algorithms such as Support Vector Machines (SVM) and Random Forests demonstrating superior performance in handling diverse datasets. The incorporation of soil parameters into these models emerges as a key innovation, as soil characteristics significantly influence land use patterns.

S. No.	Title of the Paper	Description
1	B. Bhattacharya, D.P. Solomatine, 2006, Machine learning in soil classification.	The document discusses a machine learning approach to automate the classification of soil based on measured data from Cone Penetration Testing (CPT). The proposed methodology involves two steps: segmentation and classification. In the segmentation step, a clustering algorithm called CONCC is used to identify contiguous blocks of instances. The algorithm takes into account the spatial variability of the data and aims to find segments similar to those identified by experts. In the classification step, decision trees, artificial neural networks (ANN), and support vector machines (SVM) are used to classify the segments into different soil classes. The methodology is tested on CPT data from the Nesselande area in the Netherlands, and the results show satisfactory classification accuracy.
2	Eman A. Alshari, Mohammed Abdulkareem, Bharti W. Gawali, 2023, Classification of land use/land cover using artificial intelligence (ANN-RF).	This paper focuses on the classification of land use/land cover using artificial intelligence algorithms. The study utilizes satellite images from the Landsat-8 and Sentinel-2A satellites to develop a classification model called ANN_RF, which combines artificial neural networks (ANN) with random forest (RF) algorithms. The proposed model aims to improve the accuracy of land use/land cover classification compared to individual satellite classifiers. The study includes a workflow diagram and a location map of the case study area.
3	Swapan Talukdar, Pankaj Singha, Susanta Mahato, Shahfahad, Swades Pal, Yuei-An Liou, Atiqur Rahman, 2020, Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations.	The document is of land-use land-cover (LULC) classification using machine learning classifiers for satellite observations. It compares the accuracy of six classifiers and concludes that random forest is the best for LULC mapping. The study focuses on a dynamic river island area in India and evaluates the performance of the classifiers in different LULC classes.
4	Yaren Aydın, Ümit sıkda g, Gebrail Bekda s, Sinan Melih Nigdeli, Zong Woo Geem, 2023, Use of Machine Learning Techniques in Soil Classification.	The document discusses the use of machine learning techniques in soil classification. It presents a case study using a dataset of 805 soil samples from a metro line construction project. The study compares the performance of different machine learning algorithms and highlights the potential of new gradient-boosting methods in achieving high classification accuracy rates.
5	Brandon Heung, Hung Chak Ho, Jin Zhang, Anders Knudby, Chuck E. Bulmer, Margaret G. Schmidt, 2015, An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping.	The overall paper is a condensed version of a research study or project that offers a brief and concise summary of its main components. It typically includes an introduction that outlines the research question, objectives, and importance of the study. The methodology section highlights the research design and data collection methods employed. The results section presents the main findings, while the discussion and conclusion sections interpret the results and provide a final assessment of the study's implications. The overall paper is a convenient resource for readers to quickly grasp the essence of a research study.
6	Hanan Samadi1, Jafar Hassanpour,2022, Classification of Soil Based on Laboratory Tests by Using Classification Algorithms in Machine Learning.	Soil classification is essential in tunnelling and civil engineering projects to evaluate the mechanical properties of soil. Machine learning classifiers, such as Kernel SVM and K Nearest Neighbors, can be used to classify different types of soil based on parameters like fines content, gravel content, sand content, liquid limit, and plasticity index. The study used 143 soil samples from the Line 7 metro project in Tehran to design and test the machine learning algorithms for soil classification. The developed networks showed high accuracy in classifying different types of soil, with only a few incorrectly identified samples Accurate soil classification models based on machine learning can help improve the precision of geotechnical projects and reduce financial and cost problems The use of Kernel SVM and K Nearest Neighbors algorithms in soil classification involves finding hyper-planes that differentiate between different soil classes with high precision.

7	Yaren Aydın, Ümit Işıkda, Gebrail Bekdaş, Sinan Nigdeli, Zong Woo Geem, Use of Machine Learning Techniques in Soil Classification	The provided sources discuss the use of machine learning techniques in soil classification, specifically in the context of the design of reliable structures and the minimization of time and cost in the classification process. The studies focus on the application of machine learning algorithms, such as XGBoost, LightGBM, and CatBoost, for soil classification, and the evaluation of their performance using cross-validation. The sources also highlight the challenges faced in soil classification, including missing data and class imbalance in the dataset. To address these challenges, data imputation techniques, such as KNN imputation, and resampling techniques, such as SMOTE, were applied to improve the accuracy of the classification models. The results show that the use of machine learning techniques, along with data preprocessing and handling class imbalance, can lead to significant improvements in the accuracy of soil classification models.
8	Jase Sitton, Brett Story, 2016, Estimating soil classification via quantitative and qualitative field testing for use in constructing compressed earth blocks	Machine Learning (ML) is used as an alternative approach for estimating soil information based on limited soil tests in construction projects. ML models are developed to predict and map geotechnical properties and soil types in the Saemangeum Onshore Solar Farm area in South Korea. The ML models in this study are constructed using two algorithms: Gaussian Process Regression (GPR) and K-nearest Neighbor (KNN). The ML models predict unexplored soil data, such as SPT-N (Standard Penetration Test-N) values and soil type, based on available data sets. The predicted data is then used to create three-dimensional maps of geotechnical properties and soil types in the study area. The study shows that the proposed ML models perform well in predicting and mapping unknown geotechnical properties and soil types, providing engineers with a preliminary understanding of soil conditions during the planning and construction of new structures. The KNN algorithm, one of the ML algorithms used in this study, calculates the distance between new data and existing data to determine the class or category
9	Hyeong-Joo Kim, Kevin Bagas, Arifki Mawuntu, Hyeong-Soo Kim, Tae-Woong Park, James Reyes, Jun-Young Park, Yeong-Seong Jeong, Machine learning for geotechnical properties and soil type mapping	Machine Learning (ML) is used as an alternative approach for estimating soil information based on limited soil tests in construction projects. ML models, constructed using Gaussian Process Regression (GPR) and K-nearest Neighbor (KNN) algorithms, are developed to predict and map geotechnical properties and soil types in the Saemangeum Onshore Solar Farm area in South Korea. The ML models predict unexplored soil data, such as SPT-N values and soil type, based on available data sets. The predicted data is then used to create three-dimensional maps of geotechnical properties and soil types in the study area the study shows that the proposed ML models perform well in predicting and mapping unknown geotechnical properties and soil types, providing engineers with a preliminary understanding of soil conditions during the planning and construction of new structures.
10	Sayed Moustaf, Mohamed Abdalzah, Mohamed Metwaly, Eslam Elawadi, Nassir Al-Arifi, 2021, A Quantitative Site-Specific Classification Approach Based on Affinity Propagation Clustering	The paper presents a novel non-objective and data-driven approach for preliminary seismic site-specific classification maps using machine learning based on affinity propagation (AP) clustering. The proposed model aims to overcome clustering errors due to the dependency on the interpreter's experience. Measurements of ambient vibrations were performed to cover the entire campus area, and the obtained results illustrate that the microtremor spectral ratio can be a robust tool in determining site effects. The AP clustering algorithm is benchmarked and calibrated by comparing it with the optimal number of clusters obtained from the K-means clustering algorithm.
11	Milos Kovacevica, Branislav Bajat, Bosko Gajic, 2010, Soil type classification and estimation of soil properties	Support vector machines (SVMs) are a powerful machine learning algorithm that has been successfully applied to a wide range of tasks, including soil type classification and estimation of soil properties. SVMs are particularly well-suited for these tasks because they are able to handle complex nonlinear relationships between input data and output labels. SVMs have demonstrated superior performance compared to traditional methods, as they effectively delineate boundaries between different soil types. Several studies have explored the application of SVMs in soil property estimation. These properties may include essential soil attributes like moisture content, organic carbon, and texture. SVMs exhibit robustness in modeling the non-linear relationships between soil properties and influencing factors, such as climate, land use, and topography. The use of SVMs in conjunction with remote sensing data has also been investigated, allowing for efficient and accurate prediction of soil properties over large areas.
12	Shiny Abraham, Chau Huynh and Huy Vu, 2020, Classification of Soils	Hydrologic soil groups (HSGs) are a critical tool for understanding and predicting the infiltration, runoff, and erosion potential of soils. Traditionally, HSGs have been classified using manual methods that are time-consuming and labor-intensive. Machine

	into Hydrologic Groups Using Machine Learning	learning (ML) offers a promising alternative for automating this process. The effectiveness of ML for HSG classification can be attributed to several factors. ML algorithms can handle complex relationships between soil properties and HSGs, which is challenging for traditional methods. Additionally, ML algorithms can learn from large datasets, which can improve classification accuracy. This classification is based on a combination of soil properties, including texture, structure, depth, and organic matter content.
13	P. Bhargavi, S.Jyothi, 2009, Applying Data Mining Technique for Classification of Agricultural Land Soils	The application of Naive Bayes in the classification of agricultural land soils has been a subject of interest in the field of agricultural data mining. Researchers have explored its efficacy in predicting soil types based on various soil attributes. A study demonstrated the successful utilization of Naive Bayes in predicting soil classifications with a focus on key features such as pH levels, nutrient content, and soil texture. The algorithm's simplicity and assumption of feature independence make it suitable for handling diverse soil data. The inherent probabilistic nature of Naive Bayes aligns well with the uncertainty inherent in soil classification tasks, providing a framework for handling diverse soil types. As handling continuous data and addressing potential dependencies among features have also been discussed in the literature. Researchers have proposed enhancements to the basic Naive Bayes model to address challenges and improve its applicability to real agricultural datasets. The literature underscores the potential of Naive Bayes as a valuable tool for classifying agricultural land soils, while also acknowledging the need for thoughtful feature selection and potential modifications to accommodate the complexities of soil data. Further exploration and refinement of this technique hold promise for advancing precision agriculture and sustainable land management practices.
14	Zhuan Zhao, Wenkang Feng, Jinrui Xiao, Xiaochu Liu, Shusheng Pan, and Zhongwei Liang, 2022, Rapid and Accurate Prediction of Soil Texture Using an Image-Based Deep Learning Autoencoder Convolutional Neural Network Random Forest (DLAC-CN-RF) Algorithm	At the algorithm employs an autoencoder convolutional neural network (CNN) to extract intricate features from soil images. This autoencoder serves as a powerful unsupervised learning tool, enabling the model to autonomously uncover essential patterns within the images. Subsequently, a Random Forest (RF) ensemble classifier utilizes these extracted features for rapid and accurate soil texture prediction. The amalgamation of deep learning and ensemble methods provides a compelling solution that not only accelerates the prediction process but also maintains a high level of accuracy. This research underscores the potential of the DLAC-CNN-RF algorithm in transforming the soil texture analysis, offering a sophisticated tool for precision agriculture and environmental management. The study not only contributes to the evolving field of machine learning applications in agriculture but also prompts further exploration into the adaptability and generalizability of such models across diverse soil conditions and geographic regions. Presents a study on image acquisition system for predicting soil images which clearly presents the prediction results. The soil texture is typically expressed as a percentage of sand, silt, and clay.
15	Machbah Uddin, akib Hassan, 2022, A novel feature-based algorithm for soil type classification.	Soil type has been classified using chemical and physical methods, which are time-consuming and expensive. Image-based soil classification is a more rapid and cost-effective method, but it has been challenging to achieve high accuracy with this approach. A novel feature-based algorithm for soil type classification has been proposed that combines quartile histogram-oriented gradients (Q-HOG), most frequent φ-Pixels, and a new feature selection method. The experimental results show that the performance of the proposed method is higher than that of existing image-based soil classification systems.
16	Manuel Campos-Taberner, 2020, Understanding deep learning in land use classification based on Sentinel-2 time series	This chapter discusses the functional properties of urban soils and their role in urban ecosystems. In addition to this chapter, several other studies on urban soils have been carried out. It investigated the chemical, physical, and biological properties of urban soils.
17	Shubadip Paul,2021, classification and characteristics of urban soil	These provide a detailed description of the methodology used in the study, including data collection and preprocessing steps, RNN architecture, and evaluation metrics used to compare model performance.
18	Azher Ibrahim Al-Taei,2022 Land Use/Land Cover Change Analysis Using MultiTemporal Remote	This research emphasizes the importance of continued research in this area for environmental planning and resource management. Overall, it serves as a valuable resource for understanding the advancements and applications of LULC change analysis using advanced technologies

	Sensing Data: A Case Study of Tigris and Euphrates Rivers Basi	
19	Yashon O. Ouma,2023. Urban land-use classification using machine learning classifiers: comparative evaluation and post-classification multi-feature fusion approach	In this study, they used the confusion matrix to evaluate the performance of the classifiers, further supporting the result that RF outperforms all other classifiers. This study makes a significant contribution to the field of his LULC classification of cities and shows convincing results
20	Islam Atef · Wael Ahmed Ramadan H. Abdel-Maguid, 2023, Modelling of land use land cover changes using machine learning and GIS techniques: a case study in El-Fayoum Governorate, Egypt.	The study utilized Geographic Information System (GIS) techniques and ground truth data from Google Earth images to evaluate LULC classes. The SVM procedure provided the most accurate maps compared to other methods. Overall, the study provides valuable insights into LULC changes and the impact of urban expansion in the region

3. CONCLUSION:

1. The dataset for soil classification includes parameters such as fines content, gravel content, sand content, liquid limit, and plasticity index. The input parameter depends on the classification and changes with the classification.
2. To achieve accurate bearing capacity predictions, Artificial Neural Networks (ANN) and Support Vector Machines (SVM) proved effective in automating soil classification based on parameters like Cone Penetration Test (CPT).
3. Emphasis is placed on the significance of high-quality datasets and the need for a balance between model complexity and simplicity in soil classification using machine learning.
4. Classification models demonstrated higher accuracy when tailored to specific soil types rather than using general classifications.
5. Research papers on soil classification using machine learning showcased a diverse range of algorithms, emphasizing the importance of interpretability in model outputs.
6. Synthesis of research papers provides a comprehensive understanding of the current state of soil classification through machine learning.
7. Evaluation of five machine learning algorithms (KNN, SVM, Decision Tree, Classification Bagged Ensemble, and Tree Bagger) revealed that KNN and SVM showed the highest classification accuracy.
8. The dataset was split into training (70%) and testing (30%) data, with models achieving high accuracy in classifying soil types.
9. Soil-type classification based on specific chemical and physical soil properties exhibited a high degree of linearity.
10. In the realm of soil classification analyses, Random Forest demonstrated the highest accuracy, underscoring its effectiveness in this domain.

In general, machine learning, particularly Random Forest, ANN, and SVM, proves highly effective in soil classification, emphasizing algorithm diversity and dataset quality. The synthesis of research underscores the balance between model complexity and simplicity, offering valuable insights for advancing automated soil classification, critical for engineering and agricultural applications.

Acknowledgements

The authors wish to acknowledge M/s GMR Institute of Technology for moral support.

References

- B. Bhattacharya, D.P. Solomatine, 2006, Machine learning in soil classification.
- Eman A. Alshari, Mohammed Abdulkareem, Bharti W. Gawali, 2023, Classification of land use/land cover using artificial intelligence (ANN-RF).
- Swapan Talukdar, Pankaj Singha, Susanta Mahato, Shahfahad, Swades Pal, Yuei-An Liou, Atiqur Rahman, 2020, Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations.
- Yaren Aydın, Ümit sıkda g, Gebrail Bekda s, Sinan Melih Nigdeli, Zong Woo Geem, 2023, Use of Machine Learning Techniques in Soil Classification.

- Brandon Heung, Hung Chak Ho, Jin Zhang, Anders Knudby, Chuck E. Bulmer, Margaret G. Schmidt, 2015, An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping.
- Hanan Samadi1, Jafar Hassanpour,2022, Classification of Soil Based on Laboratory Tests by UsingClassification Algorithms in Machine Learning.
- Yaren Aydın, Ümit Işıkda, Gebrail Bekdaş, Sinan Nigdeli, Zong Woo Geem, Use of Machine Learning Techniques in Soil Classification
- Jase Sitton, Brett Story, 2016, Estimating soil classification via quantitative and qualitative field testing for use in constructing compressed earth blocks
- Hyeong-Joo Kim, Kevin Bagas, Arifki Mawuntu, Hyeong-Soo Kim, Tae-Woong Park, James Reyes, Jun-Young Park, Yeong-Seong Jeong, Machine learning for geotechnical properties and soil type mapping
- Sayed Moustaf, Mohamed Abdalzah, Mohamed Metwaly, Eslam Elawadi, Nassir Al-Arifi, 2021, A Quantitative Site-Specific Classification Approach Based on Affinity Propagation Clustering
- Milos Kovacevica, Branislav Bajat, Bosko Gajic, 2010, Soil type classification and estimation of soil properties
- Shiny Abraham, Chau Huynh and Huy Vu, 2020, Classification of Soils into Hydrologic Groups Using Machine Learning
- P. Bhargavi, S.Jyothi, 2009, Applying Data Mining Technique for Classification of Agricultural Land Soils
- Zhuan Zhao, Wenkang Feng, Jinrui Xiao, Xiaochu Liu, Shusheng Pan, and Zhongwei Liang, 2022, Rapid and Accurate Prediction of Soil Texture Using an Image-Based Deep Learning Autoencoder Convolutional Neural Network Random Forest (DLAC-CN-RF) Algorithm
- Machbah Uddin, akib Hassan, 2022, A novel feature-based algorithm for soil type classification.
- Manuel Campos-Taberner, 2020, Understanding deep learning in land use classification based on Sentinel-2 time series
- Shubadip Paul,2021, classification and characteristics of urban soil
- Azher Ibrahim Al-Taei,2022 Land Use/Land Cover Change Analysis Using Multi-Temporal Remote Sensing Data: A Case Study of Tigris and Euphrates Rivers Basin
- Yashon O. Ouma,2023. Urban land-use classification using machine learning classifiers: comparative evaluation and post-classification multi-feature fusion approach
- Islam Atef · Wael Ahmed Ramadan H. Abdel-Maguid, 2023, Modelling of land use land cover changes using machine learning and GIS techniques: a case study in El-Fayoum Governorate, Egypt.